Enhancing Traffic Sign Detection with ToDayGAN and YOLOv5: A Two-Step Data Augmentation Approach for Small Datasets

Benson Zu
Computational and Mathematical Engineering
zuyeyang@stanford.edu

Abstract

Accurate traffic sign detection is critical for autonomous driving systems, but varying illumination conditions such as daytime and nighttime pose challenges to detection algorithms. Traditional methods, either feeding the model with giant datasets or using algorithmic method to remove noise , are hard to reach a balance between efficiency and effectiveness. To address this issue, we propose an integrated approach with small datasets, leveraging ToDayGAN[3] for nighttime-to-daytime image translation and YOLOv5[5] for object detection, targeting state-of-art performance. The model fine-tuned with the augmented dataset demonstrated significant improvements, with precision from 0.961 to 0.977, recall from 0.928 to 0.943, and mean Average Precision (mAP) increasing from 0.765 to 0.784 These results validate the effectiveness of our approach in addressing the illumination variability challenge in traffic sign detection.

1. Introduction

The ability to detect traffic signs accurately under varying illumination conditions is critical for the reliability and safety of autonomous driving systems. Robust detection in both daytime and nighttime conditions enhances the overall performance of these systems, contributing to safer navigation and decision-making processes. Traditional methods often rely on complex models to address noise in images captured under different conditions[6][2][4], but these approaches can be limited in their flexibility and effectiveness. Furthermore, neural network-based methods, while providing improved performance, are often designed to generalize well across multiple domains. This project aims to address these challenges by leveraging ToDayGAN[3] for nighttime-to-daytime image translation as a data augmentation technique, followed by fine-tuning YOLOv5[5] for traffic sign detection on a small dataset.

This project proposes a two-step approach to improve traffic sign detection in a small dataset. First, we use ToDayGAN to convert nighttime images to daytime images, effectively augmenting the dataset. Second, we fine-tune YOLOv5 on the augmented dataset to enhance its detection capabilities. Our experiments demonstrate that this integrated approach significantly improves detection accuracy under varying illumination conditions. The results show a notable increase in mean average precision (mAP), highlighting the potential of combining ToDayGAN and YOLOv5 for robust traffic sign detection.

2. Related Work

Traditional approaches often rely on complex mathematical models to address noise in images, such as the low-rank model for the dark channel prior for fog removal[6][2][4]. These methods, while effective, are typically designed for specific conditions and may not generalize well across multiple domains.

Generative Adversarial Networks (GANs) have shown great promise in image translation tasks, including weather condition normalization[7]. Weather GAN, for example, has demonstrated impressive results in transferring weather conditions among sunny, cloudy, foggy, rainy, and snowy states . Our approach leverages ToDayGAN, a specific type of GAN, to convert nighttime images to daytime images, thereby augmenting the dataset and addressing illumination challenges.

3. Data

We use photo datasets taken by Erik Mclean on Unsplash from Kaggle[1]. The dataset used for this project consists of 877 images, each sized ranged from 280 300 to 400 560 pixels1. It includes four distinct classes of road signs: Traffic Light, Stop, Speedlimit, and Crosswalk. Bounding box annotations are provided in the PASCAL VOC format.

All images were resized to 300*400 to ensure consistent dimensions suitable for input into neural network models. Besides, the annotations were converted to the YOLO format using a given *preprocess.py* module from the

datasets[1]. This conversion includes parsing the XML files containing the bounding box coordinates, and then normalizing the bounding box coordinates relative to the image dimensions, and saving the converted annotations in a text file format compatible with YOLO.

Figure 1: Sample Images from the Datasets

Inspecting the datasets with a summary (Figure 2, we can Speed Limit signs has the largest population(approximately 600 instances). Traffic Light and Crosswalk each have around 100 instances, while Stop signs are the least represented with about 50 instances. Bounding box annotations show a concentration of traffic signs near the image centers, typical for such datasets. The scatter plots of bounding box centers and dimensions show that most signs are centrally located and relatively small, with a consistent width-to-height ratio. This distribution represents the dataset's diversity and structure, which will be helpful for training effective object detection models.

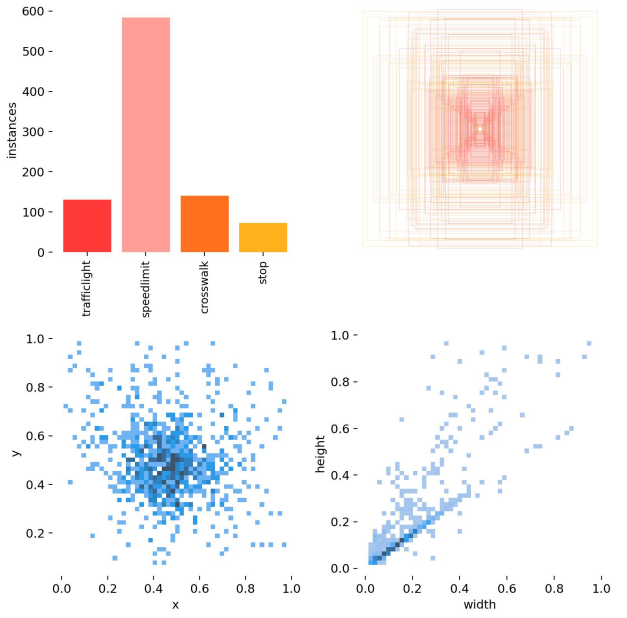

Figure 2: Instance Summaries of the Datasets including instance distribution in numbers and positions

4. Methods

4.1. ToDayGAN Architecture

We used ToDayGAN is a modified image-translation model designed to convert nighttime images to daytime representations[3]. The core components of ToDayGAN include:

4.1.1 Generator Networks

The encoder extracts features from the input image, while the decoder reconstructs the image in the target domain (daytime). The generators are trained to translate images from one domain (night) to another (day) without paired examples, using a cycle-consistent adversarial network approach.

4.1.2 Discriminator Networks

ToDayGAN[3] employs multiple discriminators to improve the quality of the translated images. Each discriminator focuses on different aspects of the input image:

- **Grayscale Discriminator**: Focuses on the luminance of the image.
- **Blurred-RGB Discriminator**: Focuses on the color and texture by using a blurred version of the image.
- **Gradient Discriminator**: Focuses on the edges and gradients of the image.

These discriminators ensure that the translated images retain essential features necessary for object detection.

4.1.3 Loss Functions

- **Adversarial Loss**: Ensures that the translated images are indistinguishable from real images in the target domain.
- **Cycle Consistency Loss**: Ensures that an image translated to the target domain and back to the source domain remains unchanged. This loss[3] is defined as:

$$L_{\text{cyc}} = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[||G_B(G_A(x)) - x||_1\right] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[||G_A(G_B(y)) - y||_1\right] \quad (1)$$

- **Relativistic Loss**: Modifies the discriminator loss[3] to improve stability and performance:

$$L_{\text{GAN}}(G, D, A, B) = \mathbb{E}_{a\sim p_{\text{data}}(a), b\sim p_{\text{data}}(b)}\left[(D_B(b) - D_B(G_A(a)) - 1)^2\right. \quad (2)$$

4.2. YOLOv5 Architecture

YOLOv5 (You Only Look Once version 5)[5] is a state-of-the-art object detection model, which is designed to predict bounding boxes and class probabilities directly from full images in a single evaluation. Key components include:

4.2.1 Grid-based Prediction

The photo will be divided into an $S \times S$ grid, in which each cell will predict its bounding boxes class, consider probability that an object's center falls in the range[5]. The confidence score reflects the likelihood of detecting an object.

4.2.2 Bounding Box Prediction

Each bounding box prediction includes coordinates (x, y) for the center, width w, height h, and a confidence score[5]. The confidence score is defined as:

$$\text{confidence} = \Pr(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}} \tag{3}$$

4.2.3 Loss Function

The loss function used during training combines localization loss (errors in bounding box coordinates), confidence loss, and classification loss. The overall loss[5] function is:

$$\begin{aligned} L = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + \\ (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \\ \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 + \\ \sum_{i=0}^{S^2} 1_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \end{aligned} \tag{4}$$

Here, λ_{coord} and λ_{noobj} are scaling factors to adjust the balance between localization and confidence errors.

4.3. Integrated Approach

The integration of ToDayGAN and YOLOv5 in this project involves the following steps:

4.3.1 Data Augmentation

ToDayGAN is used to convert all nighttime images in the dataset to daytime images, effectively doubling the dataset size and improving its diversity. This addresses the illumination condition problem by normalizing the lighting conditions across the dataset. And the converted "fake" images will share the same annotation as the original images3.

Figure 3: Examples of Converting Night Images (Left) to Fake Day Images (Right)

4.3.2 Fine-Tuning YOLOv5

YOLOv5 is fine-tuned using the augmented dataset, leveraging the increased diversity to improve its robustness and detection accuracy. The model is trained end-to-end on the augmented dataset, ensuring that it learns to detect traffic signs effectively under varied illumination conditions.

By combining ToDayGAN for image translation and YOLOv5 for object detection, this approach leverages advanced techniques in generative adversarial networks and deep learning-based object detection to achieve robust and accurate traffic sign detection in diverse lighting conditions.4

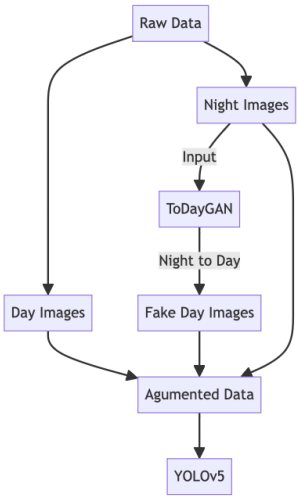

Figure 4: Method Workflow

5. Experiments

The experiments were divided into two main parts: fine-tuning YOLOv5 with the original dataset and fine-tuning YOLOv5 with the augmented dataset.

5.1. Data Splitting

The dataset was split into training, validation, and testing sets with the following proportions: Training set: 75%; Validation set: 15% ;Testing set: 10%.

5.2. Experiment 1: Fine-Tuning YOLOv5 with Original Dataset

The configuration for the YOLOv5 training usings automatic batch size by Yolo5, traning for 40 epochs with the pretrained model *yolov5s.pt* [5]

5.3. Experiment 2: Fine-Tuning YOLOv5 with Augmented Dataset

In the second experiment, we used ToDayGAN to convert nighttime images in training datasets to daytime images, thereby augmenting the dataset by adding them back to the training datasets. We then fine-tuned YOLOv5 using this augmented dataset. The configuration for the YOLOv5 training remains the same, but the dataset configuration file points to the augmented dataset.

5.4. Results and Analysis

The plots below display the training and validation metrics for YOLOv5 models trained on both the original and augmented datasets. The graphs illustrate various loss functions and performance metrics over 40 epochs. For the model using original datasets 5 and augmented dataset 6, both of their training box loss, object loss, and classification loss decrease steadily, indicating effective learning. The validation losses follow a similar trend, confirming that the model generalizes well to unseen data. For both cases, Precision, recall, and mAP (mean Average Precision) metrics improve significantly over the epochs. Precision and recall stabilize near high values, while mAP metrics demonstrate consistent improvements, indicating enhanced detection accuracy.

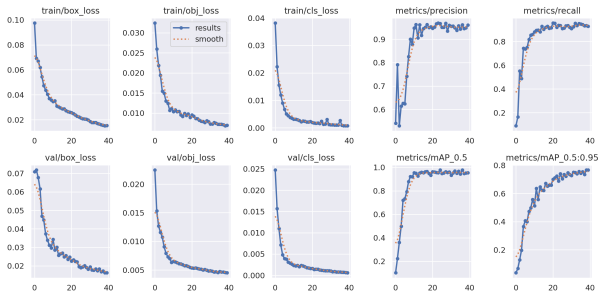

Figure 5: Summary of Losses of YOLOv5 with original datasets

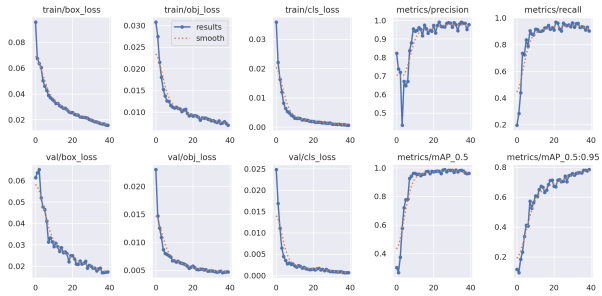

Figure 6: Summary of Losses of YOLOv5 with ToDayGAN augumented datasets

The performance of the trained YOLOv5 models was evaluated on the testing set, and the results were compared between the models trained on the original dataset and the augmented dataset. The key metrics used for evaluation included mean Average Precision (mAP), Precision, Recall.

Metric	YOLOv5 (Original)	YOLOv5 (Augmented)
Precision	0.961	0.977
Recall	0.928	0.943
mAP	0.765	0.784

Table 1: Overall Performance Comparison

Class	P	P(Aug)	mAP50	mAP50(Aug)
trafficlight	0.885	0.947	0.602	0.616
speedlimit	0.992	0.992	0.863	0.874
crosswalk	0.97	1.000	0.680	0.708
stop	0.968	0.969	0.913	0.939

Table 2: Class-wise Performance Comparison

The results indicate a significant improvement in detection performance when using the augmented dataset. The mAP increased from 0.765 to 0.784, demonstrating the effectiveness of using ToDayGAN for data augmentation. Additionally, both precision (from 0.961 to 0.977) and recall (from 0.928 to 0.943) values improved with the model using augmented datasets.

To visualize our predictions, we can use labeled images as a reference (see Figure 7) to compare predictions made by YOLOv5 with and without augmented datasets. When comparing the predictions made with the original dataset (Figure 8) to those made with the augmented dataset (Figure 9), it is evident that for bright and large targets, the predictions are almost the same. However, checking images at index positions (0,0) and (2,2), we notice that YOLOv5 with the regular dataset missed detecting the crosswalk sign in both cases. These missed targets are often hidden in the dark background at a distance, resulting in lower confidence intervals for the YOLOv5 model during detection.

With the augmentation using ToDayGAN, the fake daytime images likely emphasize the visibility of signs under low luminance conditions by enhancing their edges and gradients. This enhancement makes the model more likely to detect targets, resulting in higher recall and precision values.

Figure 7: Correct Labels

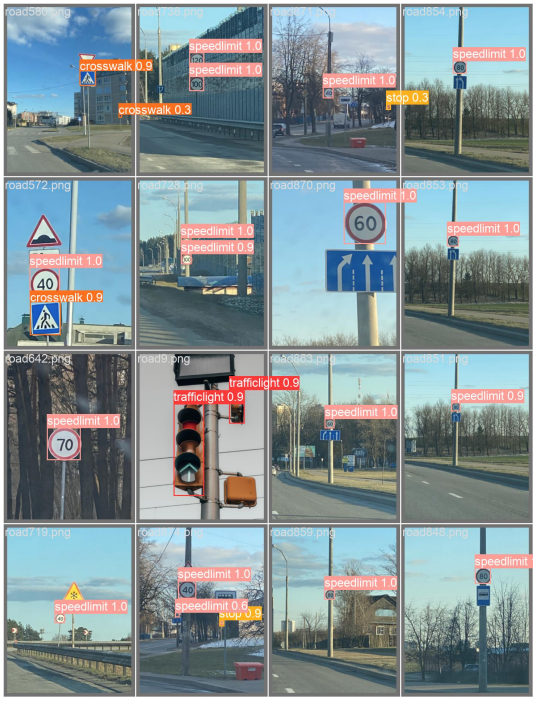

Figure 8: Prediction by YOLOv5

Figure 9: Prediction by GAN Agumented YOLOv5

This improvement is attributed to the increased diversity and robustness of the training dataset, which helps the model generalize better to different illumination conditions.

6. Conclusion

The project demonstrates the integrated approach of using ToDayGAN[3] for data augmentation and YOLOv5[5] for object detection can improve traffic sign detection under varied illumination conditions. By converting nighttime images to daytime images, ToDayGAN enhances the dataset's diversity, enabling YOLOv5 to achieve higher mean Average Precision (mAP), precision, and recall values. This improvement is attributed to the augmented dataset's ability to emphasize the visibility of traffic signs under low luminance conditions, making the model in detecting these signs more accurately.

Future work could explore further enhancements in several areas. One potential direction is to incorporate additional data augmentation techniques to address other challenging conditions such as adverse weather, including raining, foggy.etc. Another area for improvement is optimizing the YOLOv5 architecture to further boost detection performance. Extending this approach to other domains requiring robust detection under varied conditions, such as pedestrian detection at night, could also provide valuable insights and applications for enhancing autonomous driving systems.

References

[1] Road signs dataset.

[2] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and A. C. Bovik. Day and night-time dehazing by local airlight estimation. *IEEE Transactions on Image Processing*, 29:6264–6275, 2020.

[3] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool. Night-to-day image translation for retrieval-based localization. *arXiv preprint arXiv:1809.09767*, March 2019.

[4] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963. IEEE, 2009.

[5] G. Jocher, A. Chaurasia, A. Jakubowski, and K. Skalski. ultralytics/yolov5: v3.1 - bug fixes and performance improvements, Dec. 2021.

[6] L. Y. Y. Chang and S. Zhong. Transformed low-rank model for line pattern noise removal. *IEEE International Conference on Computer Vision*, 2017.

[7] H. Z. D. L. Y. X. H.-a. L. Zhang, Jing and X. Li. Research on rainy day traffic sign recognition algorithm based on pmrnet. *Mathematical Biosciences and Engineering*, 20(7), 2023.